

## General Motif Characteristics:

### Top Similar Looking Motifs:

ENCODE.UA1	0.036	
HOMER.GFX(?) .Promoter	0.038	
HOMER.ZBTB33.GM12878-ZBTB33-ChIP-Seq	0.065	
HTSELEX.ZBED1.znf_BED.DBD.dimeric.wn1	0.195	
HCMC.KAISO_f1.wn	0.215	
HCMC.TFDP1_f1.wn	0.274	
HCMC.E2F4_do.wn	0.296	
HCMC.E2F1_f2.wn	0.297	
HCMC.E2F3_si.wn	0.302	
UNIPROBE.Irf6_secondary.wn	0.304	

### Statistics:

- Motif name: denovo\_WM\_5
- Enrichment: 2.91626744534
- log-Likelihood Ratio: 535.152263225
- Area under precision recall curve: 0.5220

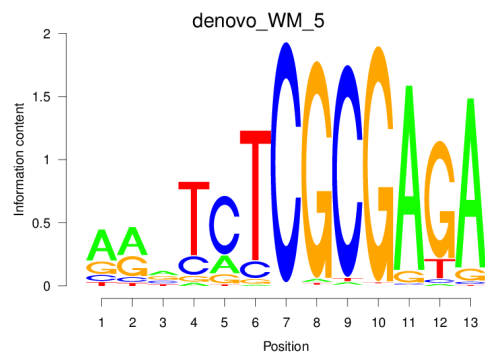


Figure 1: Logo of Weight Matrix

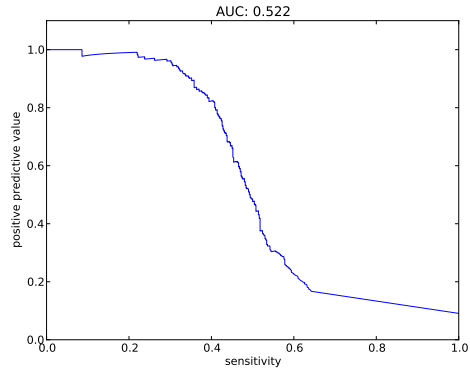


Figure 2: Precision and Recall curve

## Statistics for TFBS - Peak Centering and Peak Posterior - Peak Height Correlation:

### 0.1

Overall statistics:

- Number of true peaks out of total number of peaks: 1425/37435
- 3.81 percent are true.
- Cut-off: minimum summed posterior of 0.20
- Peak plots contain TFBS of posterior  $\geq 0.20$

Statistic for centering of TFBSs at peak centers:

- Enrichment at binding sites: 15.864

Correlation between peak Z-score and number of binding sites at peak: 0.4322

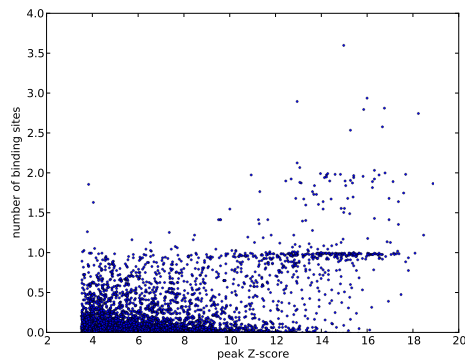


Figure 3: Scatter plot of peak Z-score versus number of binding sites at peak

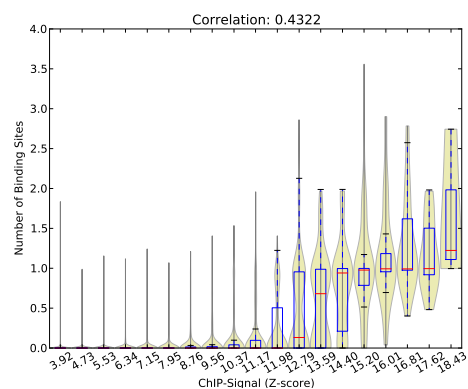


Figure 4: Violin plot of peak Z-score versus number of binding sites at peak

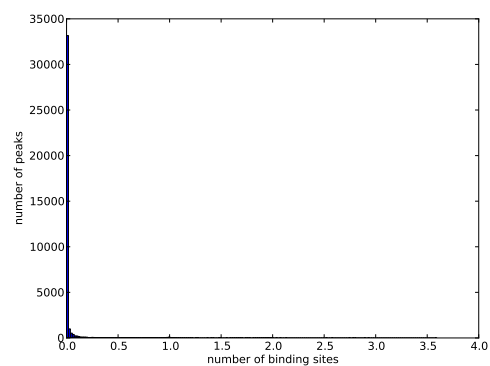


Figure 5: Histogram of number of binding sites at peaks

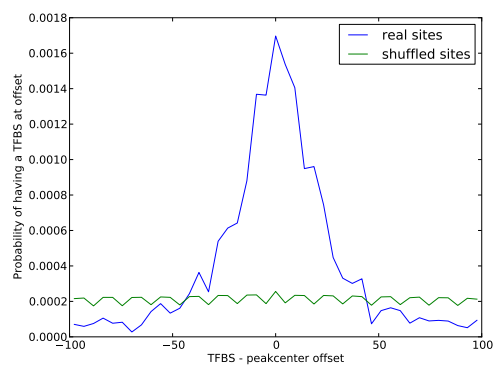


Figure 6: Histogram of distances of TFBS to its nearest peak center

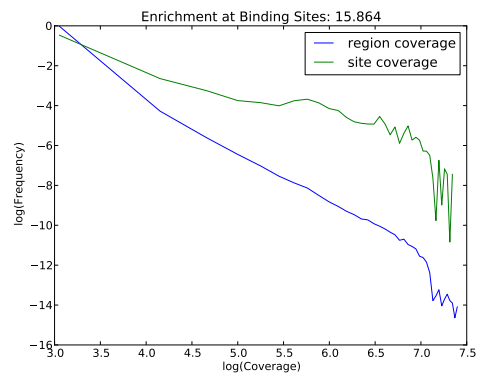


Figure 7: Histograms of coverage at sites and in total regions